Algorithmique des Réseaux Sociaux

19 Novembre

Inégalité de Cheeger

Dans tout cet exercice, G est un graphe d-régulier simple. On définit la connectivité d'une coupe par :

$$\phi(S) = \frac{e(S, V-S)}{\frac{d}{|V|}|S||V-S|},$$

et la constante isopérimétrique par : $\phi(G) = \min_{S \subset V, S \neq \emptyset, V} \phi(S)$.

L'expansion d'une coupe est définie par :

$$h(S) = \frac{e(S, V-S)}{d \min\{|S|, |V-S|\}},$$

et le taux d'expansion par : $h(G) = \min_{S \subset V, S \neq \emptyset, V} h(S)$. Le calcul de $h(G)$ est NP-difficile et le meilleur algorithme dû à Arora, Rao & Vazirani (2009) donne une $O(\sqrt{\log n})$ approximation.

1. Montrer que

$$h(G) \leq \phi(G) \leq 2h(G).$$

On considère $M = \frac{1}{d}A$ la matrice d'adjacence normalisée du graphe G de valeurs propres $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Le but de l'exercice est de montrer que

$$\frac{1-\lambda_2}{2} \leq h(G) \leq \sqrt{2(1-\lambda_2)}.$$

2. Montrer que :

$$\begin{aligned} \lambda_2 &= \sup_{\mathbf{x} \in \mathbb{R}^n,\, \|\mathbf{x}\|=1,\, \mathbf{x} \perp \mathbf{1}} \mathbf{x}^t M \mathbf{x}, \\ \lambda_n &= \inf_{\mathbf{x} \in \mathbb{R}^n,\, \|\mathbf{x}\|=1} \mathbf{x}^t M \mathbf{x} \end{aligned}$$

3. Montrer que pour tout vecteur $\mathbf{x}$,

$$\sum_{i,j} M_{ij}(x_i - x_j)^2 = 2\mathbf{x}^t\mathbf{x} - 2\mathbf{x}^t M \mathbf{x}.$$

   En déduire que

$$1 - \lambda_2 = \inf_{\mathbf{x} \in \mathbb{R}^n,\, \|\mathbf{x}\|=1,\, \mathbf{x} \perp \mathbf{1}} \frac{1}{2} \sum_{ij} M_{ij}(x_i - x_j)^2,$$

   puis que $\lambda_2 = 1$ si et seulement si G n'est pas connecté.

4. Montrer que pour tout vecteur $\mathbf{x}$:
$$\sum_{ij} M_{ij}(x_i + x_j)^2 = 2\mathbf{x}^t\mathbf{x} + 2\mathbf{x}^t M\mathbf{x}.$$
En déduire que $\lambda_n = -1$ si et seulement si une composante de G est bipartie.
5. Montrer que
$$\phi(G) = \min_{\mathbf{x}\in\{0,1\}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{ij} M_{ij}|x_i - x_j|^2}{\frac{1}{n}\sum_{ij}|x_i - x_j|^2}$$
6. Montrer que
$$1 - \lambda_2 = \min_{\mathbf{x}\in\mathbb{R}^V - \mathbb{R}\mathbf{1}} \frac{\sum_{ij} M_{ij}|x_i - x_j|^2}{\frac{1}{n}\sum_{ij}|x_i - x_j|^2}.$$
7. En déduire que $1 - \lambda_2 \le \phi(G) \le 2h(G)$.

Pour montrer l'autre inégalité, nous introduisons l'algorithme suivant :

Partition Spectrale

— Entrée : graphe $G = (V, E)$ et un vecteur $\mathbf{x} \in \mathbb{R}^V$.
— Trier les sommets de V selon l'ordre décroissant des entrées de $\mathbf{x}$, c'est à dire $V = \{v_1, \dots, v_n\}$ avec $x_{v_1} \le x_{v_2} \le \cdots \le x_{v_n}$.
— soit $i \in \{1, \dots, n-1\}$ tel que $h(\{v_1, \dots, v_i\})$ soit minimal.
— Sortie : $S = \{v_1, \dots, v_i\}$.

Etant donné un graphe G et un vecteur $\mathbf{x} \in \mathbb{R}^V$, on définit :
$$\delta = \frac{\sum_{i,j} M_{ij}|x_i - x_j|^2}{\frac{1}{n}\sum_{i,j}|x_i - x_j|^2},$$
où M est la matrice d'adjacence normalisée. Nous allons montrer que si S est la sortie correspondante de l'algorithme alors $h(S) \le \sqrt{2\delta}$.

Pour simplifier les notations, on suppose que $V = \{1, \dots, n\}$ et que $x_1 \le x_2 \le \cdots \le x_n$, de telle sorte que notre but est de montrer qu'il existe i tel que $h(\{1, \dots, i\}) \le \sqrt{2\delta}$. Pour cela nous allons utiliser la méthode probabiliste.

8. Montrer que l'on peut supposer que $x_{\lfloor n/2 \rfloor} = 0$ et $x_1^2 + x_n^2 = 1$ sans perte de généralité.
9. Soit T une variable aléatoire à valeur dans $[x_1, x_n]$ telle que $\mathbb{P}(a \le t \le b) = \int_a^b 2|t|dt$ pour $x_1 \le a \le b \le x_n$. Soit $S_T = \{i,\ x_i \le T\}$. Montrer que
$$\mathbb{E}[\min\{|S_T|, |V - S_T|\}] = \sum_i x_i^2.$$
10. Montrer que
$$\mathbb{P}((i,j) \text{ est coupé par} (S_T, V - S_T)) \le |x_i - x_j| \left(|x_i| + |x_j|\right).$$
11. En déduire que
$$\frac{1}{d}\mathbb{E}\left[e(S_T, V - S_T)\right] \le \frac{1}{2}\sqrt{\sum_{i,j} M_{ij}(x_i - x_j)^2}\sqrt{\sum_{i,j} M_{ij}\left(|x_i| + |x_j|\right)^2}$$
12. Montrer que $\sum_{i,j} M_{ij}(x_i - x_j)^2 \le 2\delta \sum_i x_i^2$ et que $\sum_{i,j} M_{ij}\left(|x_i| + |x_j|\right)^2 \le 4\sum_i x_i^2$. En déduire que $\frac{1}{d}\mathbb{E}\left[e(S_T, V - S_T)\right] \le \sqrt{2\delta}\sum_i x_i^2$.
13. En déduire qu'il existe S de la forme $\{1, \dots, i\}$ tel que $h(S) \le \sqrt{2\delta}$.
14. En déduire que $h(G) \le \sqrt{2(1 - \lambda_2)}$ et que l'algorithme *Partition Spectrale* permet de trouver S tel que $h(S) \le 2\sqrt{h(G)}$.